09/423093

58/PRTS

WO 98/50531 PCT/AU98/00315

420 Rec'd PCT/PTO 0 1 NOV 1999

Nucleic acid molecules specific for bacterial antigens and uses thereof.

TECHNICAL FIELD

The invention relates to novel nucleotide sequences located in a gene cluster which controls the synthesis of a bacterial polysaccharide antigen, especially an O antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular polysaccharide antigens (particularly O antigens) and for the identification of the polysaccharide antigens (particularly O antigens) of those bacteria.

BACKGROUND ART

Enteropathogenic E. coli strains are well known causes of diarrhoea and haemorrhagic colitis in humans and can lead to potentially life threatening sequelae including haemolytic uremic syndrome and thrombotic thrombocytopaenic purpura. Some of these strains are commonly found in livestock and infection in humans is usually a consequence of consumption of contaminated meat or dairy products which have been improperly processed. The O specific polysaccharide component (the "O antigen") of lipopolysaccharide is known to be a major virulence factor of enteropathogenic E. coli strains.

The E. coli O antigen is highly polymorphic and 166 different forms of the antigen have been defined; Ewing, W. H. [in Edwards and Ewings "Identification of the Enterobacteriacea" Elsevier. Amsterdam (1986)] discusses 128 different O antigens while Lior H. (1994) extends the number to 166 [in "Classification of *Escherichia coli* In *Escherichia coli in domestic animals and humans* pp31-72. Edited by C.L.Gyles CAB International]. The genus Salmonella enterica has 46 known O antigen types [Popoff M.Y. *et al* (1992) " Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France].

An important step in determining the biosynthesis of O antigens and therefore the mechanism of the polymorphism has been to characterise the gene clusters controlling O antigen biosynthesis. The genes specific for the synthesis of the O antigen are generally located in a gene cluster at map position 45 minutes on the chromosome of E. coli K-12 [Bachmann, B. J. 1990 "Linkage map of Escherichia coli K-12". *Microbiol. Rev.* 54: 130-197], and at the corresponding position in S. enterica LT2 [Sanderson et al (1995) "Genetic map of Salmonella enterica typhimurium", Edition VIII Microbiol. Rev. **59:** 241-303]. In both cases the O antigen gene cluster is close to the *gnd* gene as is the case in other strains of E. coli and S. enterica [Reeves P.R. (1994) "Biosynthesis and assemby of lipopolysaccharide, 281-314. in A. Neuberger and L.L.M. van Deenen (eds) "Bacterial cell wall, new comprehensive biochemistry " vol 27 Elsevier Science Publishers]. These genes encode enzymes for the synthesis of nucleotide diphosphate sugars and for assembly of the sugars into oligosaccharide units and in general for polymerisation to O antigen.

The E. coli O antigen gene clusters for a wide range of E. coli O antigens have been cloned but the O7, O9, O16 and O111 O antigens have been studied in more detail with only O9 and O16 having been fully characterised with regard to nucleotide sequence to date [Kido N., Torgov V.I., Sugiyama T., Uchiya K., Sugihara H., Komatsu T., Kato N. & Jann K. (1995) "Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 rfb gene cluster, characterisation of mannosyl transferases, and evidence for an ATP-binding cassette transport system" J. of Bacteriol. *177* 2178-2187; Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M., Redmond J.W., Lindquist L. & Reeves PR (1994) "Structure of the O antigen of *E. coli* K12 and the sequence of its *rfb* gene cluster" J. of Bacteriol. **176** 4144-4156; Jayaratne, P. *et al.* (1991) "Cloning and analysis of duplicated *rfbM* and *rfbK* genes involved in the

formation of GDP-mannose in *Escherichia coli* O9:K30 and participation of *rfb* genes in the synthesis of the group 1 K30 capsular polysaccharide" *J. Bacteriol.* 176: 3126-3139; Valvano, M. A. and Crosa, J. H.(1989)" Molecular cloning and expression in *Escherichia coli K-12* of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E.coli* O7:K1". *Inf and Immun.* 57:937-943; Marolda C. L. And Valvano, M. A. (1993). "Identification, expression, and DNA sequence of the GDP-mannose biosynthesis genes encoded by the O7 *rfb* gene cluster of strain VW187 (*Eschericia coli* O7:K1*)". J. Bacteriol*. 175:148-158.]

Bastin D.A., *et al.* 1991 ["Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E.coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231] and Bastin D.A. and Reeves, P.R. [(1995)" Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". *Gene* 164: 17-23] isolated chromosomal DNA encoding the E. coli O111 *rfb* region and characterised a 6962 bp fragment of E. coli O111 rfb. Six open reading frames (orfs) were identified in the 6962 bp partial fragment and the alignment of the sequences of these orfs revealed homology with genes of the GDP-mannose pathway, *rfbK* and *rfbM*, and other *rfb* and *cps* genes.

The nucleotide sequences of the loci which control expression of Salmonella enterica B, A, D1, D2, D3, C1, C2 and E O antigens have been characterised [Brown, P. K., L. K. Romana and P. R. Reeves (1991) "Cloning of the *rfb* gene cluster of a group C2 Salmonella enterica: comparison with the *rfb* regions of groups B and D Mol. Microbiol. **5**:1873-1881; Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of Salmonella enterica serovar typhimurium (LT2)". Mol. Microbiol. **5**:692-713; Lee, S. J., L. K. Romana, and P. R. Reeves (1992) "Sequences and structural analysis of the *rfb* (O antigen)gene cluster from a group C1 Salmonella enterica

enterica strain" J. Gen. Microbiol. **138**: 1843-1855; Lui, D., N. K. Verma, L. K. Romana, and P. R. Reeves (1991) "Relationship among the *rfb* regions of Salmonella enterica serovars A, B and D" J. Bacteriol. **173:** 4814-4819; Verma, N. K., and P. Reeves (1989) "Identification and sequence of *rfb*S and *rfb*E, which determine the antigenic specificity of group A and group D Salmonella entericae" J. Bacteriol. **171:** 5694-5701; Wang, L., L. K. Romana, and P. R. Reeves (1992) "Molecular analysis of a Salmonella enterica enterica group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism" Genetics **130:** 429-443; Wyk, P., and P. Reeves (1989). "Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B Salmonella entericae: homology with galactose epimerase" J. Bacteriol. **171:** 5687-5693,; Xiang, S. H., M. Hobbs, and P. R. Reeves. 1994 Molecular analysis of the rfb gene luster of a group D2 Salmonella enterica strain: evidence for its origin from an insertion sequence -mediated recombination event between group E and D1 strains. J. Bacteriol. 176: 4357 -4365; Curd, H., D. Liu and P. R. Reeves, 1998. Relationships among the O antigen Salmonella enterica groups B, D1, D2, and D3. J. Bacteriol. 180: 1002-1007.].

Of the closely related Shigella (which really can be considered to be part of E. coli) S. dysenteriae and S. flexneri O antigens have been fully sequenced and are next to *gnd*. [Klena JD & Schnaitman CA (1993) "Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1" Mol. Microbiol. **9** 393-402; Morona R., Mavris M., Fallarino A. & Manning P. (1994) "Characterisation of the *rfc* region of *Shigella flexneri*" J.Bacteriol **176**: 733-747]

Inasmuch as the O antigen of enteropathogenic E. coli strains and the O antigen of Salmonella enterica strains are major virulence factors and are highly polymorphic, there is a real need to develop highly specific, sensitive, rapid and inexpensive diagnostic assays to

detect E. coli and assays to detect S. enterica. There is also a real need to develop diagnostic assays to identify the O antigens of E. coli strains and assays to identify the O antigens of S. enterica strains. With regard to the detection of E. coli these needs extend beyond EHEC (enteropathogenic haemorrhagic E. coli) strains but this is the area of greatest need. There is interest in diagnostics for ETEC (enterotoxigenic E. coli) etc in E. coli.

The first diagnostic systems employed in this field used large panels of antisera raised against E. coli O antigen expressing strains or S. enterica O antigen expressing strains. This technology has inherent difficulties associated with the preparation, storage and usage of the reagents, as well as the time required to achieve a meaningful diagnostic result.

Nucleotide sequences derived from the O antigen gene clusters of S. enterica strains have been used to determine S. enterica O antigens in a PCR assay [Luk, J.M.C. *et al.* (1993) "Selective amplification of abequose and paratose synthase genes (*rfb*) by polymerase chain reaction for identification of S. enterica major serogoups (A, B, C2, and D)", *J. Clin. Microbiol.* 31:2118-2123]. The prior complete nucleotide sequence characterisation of the entire *rfb* locus of serovars Typhimurium, Paratyphi A, Typhi, Muenchen, and Anatum; representing groups B, A, D1, C2 and E1 respectively enabled Luk *et al.* to select oligonucleotide primers specific for those serogroups. Thus the approach of Luk *et al.* was based on aligning known nucleotide sequences corresponding to CDP-abequose and CDP-paratose synthesis genes within the O antigen regions of S. enterica serogroups E1, D1, A, B and C2 and exploiting the observed nucleotide sequence differences in order to identify serotype-specific oligonucleotides.

In an attempt to determine the O antigen serotype of a Shiga-like toxin producing E. coli strain, Paton, A. W., *et al.* 1996 ["Molecular microbiological investigation of an outbreak of Hemolytic-Uremic Syndrome caused by dry

fermented sausage contaminated with Shiga-like toxin producing Escherichia coli". *J. Clin. Microbiol.* **34:** 1622-1627], used oligonucleotides derived from the *wbdI* (*orf*6) region, which were believed to be specific to the E. coli O111 antigen and which were derived from E. coli O111 sequence, in a PCR diagnostic assay. Unpublished reports indicate that the approach of Paton *et al.* is deficient in that the nucleotide sequences derived from *wbdI* may not specifically identify the O111 antigen and in fact lead to detection of false positive results. Paton *et al.* disclose the detection of 5 O111 antigen isolates by PCR when in fact from only 3 of those isolates did they detect bacteria which reacted with O111 specific antiserum.

DESCRIPTION OF THE INVENTION

Whilst not wanting to be held to a particular hypothesis, the present inventors now believe that the reported false positives found with the Paton *et al.* method are due to the fact that the nucleic acid molecules employed by Paton *et al.* were derived from genes which have a putative function as a sugar pathway gene, [Bastin D.A. and Reeves, P.R. (1995) Sequence and analysis of the O antigen gene(*rfb*) cluster of Escherichia coli O111. *Gene* 164: 17-23] which they now believe to lack the necessary nucleotide sequence specificity to identify the E. coli O antigen. The inventors now believe that many of the nucleic acid molecules derived from sugar pathway genes expressed in S. enterica or other enterobacteria are also likely to lack the necessary nucleotide sequence specificity to identify specific O antigens or specific serotypes.

In this regard it is important to note that the genes for the synthesis of a polysaccharide antigen include those related to the synthesis of the sugars present in the antigen (sugar pathway genes) and those related to the manipulation of those sugars to form the polysaccharide. The present invention is predominantly concerned with the latter group of genes, particularly the assembly and

transport genes such as transferase, polymerase and flippase genes.

The present inventors have surprisingly found that the use of nucleic acid molecules derived from particular assembly and transport genes, particularly transferase, wzx and wzy genes, within O antigen gene clusters can improve the specificity of the detection and identification of O antigens. The present inventors believe that the invention is not necessarily limited to the detection of the particular O antigens which are encoded by the nucleic acid molecules exemplified herein, but has broad application for the detection of bacteria which express an O antigen and the identification of O antigens in general. Further because of the similarities between the gene clusters involved in the synthesis of O antigens and other polymorphic polysaccharide antigens, such as bacterial capsular antigens, the inventors believe that the methods and molecules of the present invention are also applicable to these other polysaccharide antigens.

Accordingly, in one aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection and identification of specific bacterial polysaccharide antigens.

The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene, *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen,
wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen.

Polysaccharide antigens, such as capsular antigens of E. coli (Type I and Type II), the Virulence capsule of S. enterica sv. Typhi and the capsules of species such as Streptococcus pneumoniae and Staphylococcus albus are

encoded by genes which include nucleotide sugar pathway genes, sugar transferase genes and genes for the transport and processing of the polysaccharide or oligosaccharide unit. In some cases these are *wzx* or *wzy* but in other cases they are quite different because a different processing pathway is used. Examples of other gene clusters include the gene clusters for an extracellular polysaccharide of Streptococcus thermophilus, an exopolysaccharide of Rhizobium melilotti and the K2 capsule of Klebsiella pneumoniae. These all have genes which by experimental analysis, comparison of nucleotide sequence or predicted protein structure, can be seen to include nucleotide sugar pathway genes, sugar transferase genes and genes for oligosaccharide or polysaccharide processing.

In the case of the E. coli K-12 colanic acid capsule gene cluster [Stevenson *et al* (1996) "Organization of the *Escherichia coli K-12* gene cluster responsible for production of the extracellular polysaccharide colanic acid". J. Bacteriol **178**: 4885-4893] genes from the three classes were identified either provisionally or definitively. Colanic acid capsule is classified with the Type I capsule of E. coli.

The present inventors believe that, in general, transferase genes and genes for oligosaccharide processing will be more specific for a given capsule than the genes coding for the nucleotide sugar synthetic pathways as most sugars present in such capsules occur in the capsules of different serotypes. Thus the nucleotide sugar synthesis pathway genes could now be predicted to be common to more than one capsule type.

As elaborated below the present inventors recognise that there may be polysaccharide antigen gene clusters which share transferase genes and/or genes for oligosaccharide or polysaccharide processing so that completely random selection of nucleotide sequences from within these genes may still lead to cross-reaction; an example with respect to capsular antigens is provided by

the E. coli type II capsules for which only transferase genes are sufficiently specific. However, the present inventors in light of their current results nonetheless consider the transferase genes or genes controlling oligosaccharide or polysaccharide processing to be superior targets for nucleotide sequence selection for the specific detection and characterisation of polysaccharide antigen types. Thus where there is similarity between particular genes, selection of nucleotide sequences from within other transferase genes or genes for oligosaccharide or polysaccharide processing from within the relevant gene cluster will still provide specificity, or alternatively the use of combinations of nucleotide sequences will provide the desired specificity. The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

Thus the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes; wherein the combination of genes is specific to the synthesis of a particular bacterial polysaccharide antigen and wherein the panel of nucleic acid molecules is specific to a bacterial polysaccharide antigen. In another preferred form, the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, together with nucleic acid molecules derived from pathway genes.

In a second aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection of bacteria which express O antigens and for the identification of the O antigens of those bacteria in diagnostic assays.

The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene, the gene being involved in the synthesis of a particular bacterial O antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial O antigen.

The nucleic acids of the invention may be variable in length. In one embodiment they are from about 10 to about 20 nucleotides in length.

In one preferred embodiment, the invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit including a *wzx* or *wzy* gene the gene being involved in the synthesis of an O antigen expressed by E. coli, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

In one more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O111 antigen (SEQ ID NO:1). More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of SEQ ID NO:1), *wzx* (nucleotide position 8646 to 9911 of SEQ ID NO:1), *wzy* (nucleotide position 9901 to 10953 of SEQ ID NO:1), *wbdM* (nucleotide position 11821 to 12945 of SEQ ID NO:1) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Table 5 and 5A, with respect to the above mentioned genes.

In another more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O157 antigen (SEQ ID NO:2). More preferably the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of SEQ ID NO:2), *wbdO*, (nucleotide position 2011 to 2757 of SEQ ID NO:2), *wbdP* (nucleotide position 5257 to

6471 of SEQ ID NO:2)), *wbdR* (13156 to 13821 of SEQ ID NO:2), *wzx* (nucleotide position 2744 to 4135 of SEQ ID NO:2) and *wzy* (nucleotide position 858 to 2042 of SEQ ID NO:2). Particularly preferred nucleic acid molecules are those set out in Table 6 and 6A.

The invention also provides in a further preferred embodiment a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit including a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by Salmonella enterica, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

In one more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica C2 antigen (SEQ ID NO:3). More preferably the sequence of the nucleic acid molecule is derived from a gene selected from the group consisting of *wbaR* (nucleotide position 2352 to 3314 of SEQ ID NO:3), *wbaL* (nucleotide position 3361 to 3875 of SEQ ID NO:3), *wbaQ* (nucleotide position 3977 to 5020 of SEQ ID NO:3), *wbaW* (nucleotide position 6313 to 7323 of SEQ ID NO:3), *wbaZ* (nucleotide position 7310 to 8467 of SEQ ID NO:3), *wzx* (nucleotide position 1019 to 2359 of SEQ ID NO:3)and *wzy* (nucleotide position 5114 to 6313 of SEQ ID NO:3). Particularly preferred nucleic acid molecules are those set out in Table 7.

In another more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica B antigen (SEQ ID NO:4). More preferably the sequence is derived from *wzx* (nucleotide position 12762 to 14054 of SEQ ID NO:4) or *wbaV* (nucleotide position 14059 to 15060 of SEQ ID NO:4). Particularly preferred nucleic acid molecules are those set out in Table 8 which are derived from *wzx* and *wbaV* genes.

In a further more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to

the S. enterica D3 O antigen and is derived from the *wzy* gene.

In yet a further preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the S. enterica E1 O antigen and is derived from the *wzx* gene.

While transferase genes, or genes coding for the transport or processing of a polysaccharide or oligosaccharide unit, such as a *wzx* or *wzy* gene, are superior targets for specific detection of individual O antigen types there may well be individual genes or parts of them within this group that can be demonstrated to be the same or closely related between different O antigen types such that cross-reactions can occur. Cross reactions should be avoided by the selection of a different target within the group or the use of multiple targets within the group.

Further, it is recognised that there are cases where O antigen gene clusters have arisen from recombination of at least two strains such that the unique O antigen type is provided by a combination of gene products shared with at least two other O antigen types. The recognised example of this phenomenon is the S. enterica O antigen serotype D2 which has genes from D1 and E1 but none unique to D2. In these circumstances the detection of the O antigen type can still be achieved in accordance with the invention, but requires the use of a combination of nucleic acid molecules to detect a specific combination of genes that exists only in that particular O antigen gene cluster.

Thus, the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. More preferably,

the panel of nucleic acid molecules is specific to the D2 O antigen and is derived from the E1 *wzy* gene and the D1 *wzx* gene.

The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes, together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

Thus, the invention also provides a panel of nucleic acid molecules, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, and sugar pathway genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the O antigen is expressed S. enterica.

Further it is recognised that there may be instances where spurious hybridisation will arise through initial selection of a sequence found in many different genes but this is typically recognisable by, for instance, comparison of band sizes against controls in PCR gels, and an alternative sequence can be selected.

The present inventors believe that based on the teachings of the present invention and available information concerning polysaccharide antigen gene clusters (including O antigen gene clusters), and through use of experimental analysis, comparison of nucleic acid sequences or predicted protein structures, nucleic acid molecules in accordance with the invention can be readily derived for any particular polysaccharide antigen of interest. Suitable bacterial strains can typically be acquired commercially from depositary institutions.

As mentioned above there are currently 166 defined E. coli O antigens while the S. enterica has 46 known O antigen types [Popoff M.Y. *et al* (1992) "Antigenic formulas of the Salmonella serovars" 6th revision WHO Collaborating centre for Reference and Research on Salmonella, Institut Pasteur Paris France]. Many other genera of bacteria are known to have O antigens and these

include Citrobacter, Shigella, Yersinia, Plesiomonas, Vibrio and Proteus.

Samples of the 166 different E. coli O antigen serotypes are available from Statens Serum Institut, Copenhagen, Denmark.

The 46 S. enterica serotypes are available from Institute of Medical and Veterinary Science, Adelaide, Australia.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more bacterial polysaccharide antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

Where a single specific oligonucleotide molecule is unavailable a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial polysaccharide. The panel of nucleic acid molecules can include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more

bacterial polysaccharide antigens comprising contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

The pair of oligonucleotide molecules may both hybridise to the same gene or to different genes. Only one oligonucleotide molecule of the pair need hybridise specifically to sequence specific for the particular antigen type. The other molecule can hybridise to a non-specific region.

Where the particular polysaccharide antigen gene cluster has arisen through recombination, the at least one pair of oligonucleotide molecules may be selected to be capable of hybridising to a specific combination of genes in the cluster specific to that polysaccharide antigen, or multiple pairs may be selected to provide hybridisation to the specific combination of genes. Even where all the genes in a particular cluster are unique, the method may be carried out using nucleotide molecules which recognise a combination of genes within the cluster.

Thus the invention provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is

specific to a particular bacterial polysaccharide antigen. The panel of nucleic acid molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are <u>E. coli</u> or <u>S. enterica</u>. More preferably, the <u>E. coli</u> express the O157 serotype or the O111 serotype. More preferably the <u>S. enterica</u> express the C2 or B serotype. Preferably, the method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the nucleic acid molecule is detected by autoradiography or detection of fluorescence.

The inventors envisage circumstances where a single specific oligonucleotide molecule is unavailable. In these circumstances a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is

expressed by S. enterica. The panel of nucleic acid molecules can include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the O111 or the O157 serotype. More preferably the S. enterica express the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis. Preferred oligonucleotides for use with O111 which provide for specific detection of O111 are illustrated in Table 5 and 5A with respect to the genes *wbdH*, *wzx*, *wzy* and *wbdM*. Preferred oligonucleotide molecules for use with O157 which provide for specific detection of O157 are illustrated in Table 6 and 6A.

With respect to serotypes C2 and B, suitable oligonucleotide molecules can be selected from appropriate regions described in column 3 of Tables 7 and 8.

The inventors envisage rare circumstances whereby two genetically similar gene clusters encoding serologically

different O antigens have arisen through recombination of genes or mutation so as to generate polymorphic variants. In these circumstances multiple pairs of oligonucleotides may be selected to provide hybridisation to the specific combination of genes. The invention thus provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. The panel of nucleic acid molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method for testing a food derived sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the

method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect the present invention relates to a method for testing a faecal derived sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one of said genes of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are <u>E. coli</u> or <u>S. enterica</u>. More preferably, the <u>E. coli</u> are of the 0111 or 0157 serotype. More preferably, the <u>S. enterica</u> are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect, the present invention relates to a method for testing a sample derived from a patient for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein

said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably, the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In the above described methods it will be understood that where pairs of oligonucleotides are used one of the oligonucleotide sequences may hybridise to a sequence that is not from a transferase, wzx or wzy gene. Further where both hybridise to one of these gene products they may hybridise to the same or a different one of these genes.

In addition it will be understood that where cross reactivity is an issue a combination of oligonucleotides may be chosen to detect a combination of genes to provide specificity.

The invention further relates to a diagnostic kit which can be used for the detection of bacteria which express bacterial polysaccharide antigens and the identification of the bacterial polysaccharide type of those bacteria.

Thus in a further aspect, the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, including a *wzx* or wzy gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide. The kit may also provide in the same or a

separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, including a *wzx* or *wzy* gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

In a further aspect the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of a bacterial O antigen. The kit may also provide in the same or a separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase , or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of O antigen, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule. Preferably the first and second nucleic acid sequences are derived from E. coli or the first and second nucleic acid sequences are derived from S. enterica.

The present inventors provide full length sequence of the O157 gene cluster for the first time and recognise that from this sequence of this previously uncloned full gene cluster appropriate recombinant molecules can be generated and inserted for expression to provide expressed O157 antigens useful in applications such as vaccines.

DEFINITIONS

The phrase, "a nucleic acid molecule derived from a gene" means that the nucleic acid molecule has a

nucleotide sequence which is either identical or substantially similar to all or part of the identified gene. Thus a nucleic acid molecule derived from a gene can be a molecule which is isolated from the identified gene by physical separation from that gene, or a molecule which is artificially synthesised and has a nucleotide sequence which is either identical to or substantially similar to all or part of the identified gene. While some workers consider only the DNA strand with the same sequence as the mRNA transcribed from the gene, here either strand is intended.

Transferase genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that transfer monomeric sugar units.

Flippase or *wzx* genes are regions of nucleic acid which have a nucleotide sequence which encodes a gene product that flips oligosaccharide repeat units generally composed of three to six monomeric sugar units to the external surface of the membrane.

Polymerase or *wzy* genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that polymerise repeating oligosaccharide units generally composed of 3-6 monomeric sugar units.

The nucleotide sequences provided in this specification are described in the sequence listing as anti-sense sequences. This term is used in the same manner as it is used in Glossary of Biochemistry and Molecular Biology Revised Edition, David M. Glick, 1997 Portland Press Ltd., London on page 11 where the term is described as referring to one of the two strands of double-stranded DNA usually that which has the same sequence as the mRNA. We use it to describe this strand which has the same sequence as the mRNA.

NOMENCLATURE

Synonyms for E. coli O111 *rfb*

Current names	Our names	Bastin et al. 1991
wbdH	orf1	
gmd	orf2	
wbdI	orf3	orf3.4*
manC	orf4	rfbM*
manB	orf5	rfbK*
wbdJ	orf6	orf6.7*
wbdK	orf7	orf7.7*
wzx	orf8	orf8.9 and rfbX*
wzy	orf9	
wbdL	orf10	
wbdM	orf11	

* Nomenclature according to Bastin D.A., *et al*. 1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol*. 5:9 2223-2231.

Other Synonyms

wzy	rfc
wzx	rfbX
rmlA	rfbA
rmlB	rfbB
rmlC	rfbC
rmlD	rfbD
glf	orf6*
wbbI	orf3#, orf8* of E. coli K-12
wbbJ	orf2#, orf9* of E. coli K-12
wbbK	orf1#, orf10* of E. coli K-12
wbbL	orf5#, orf 11* of E. coli K-12

\# Nomenclature according to Yao, Z. And M. A. Valvano 1994. "Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Eschericia coli K-12 W3110: identification of genes the confer groups-specificty to Shigella flexineri serotypes Y and 4a". J. Bacteriol. 176: 4133-4143.

* Nomenclature according to Stevenson et al. 1994. "Structure of the O-antigen of E. coli K-12 and the sequence of its rfb gene cluster". J. Bacteriol 176: 4144-4156.

- S. enterica is a name introduced in 1987 to replace the many other names such as Salmonella typhi and Salmonella typhimurium, the old species names becoming serovar names as in S. enterica sv Typhi. However, the traditional names are still widely used.
- The O antigen genes of many species were given rfb names (rfbA etc) and the O antigen gene cluster was often referred to as the rfb cluster. There are now new names for the rfb genes as shown in the table. Both terminologies have been used herein, depending on the source of the information.

- **BRIEF DESCRIPTION OF DRAWINGS**

Figure 1 shows *Eco* R1 restriction maps of cosmid clones pPR1054, pPR1055, pPR1056, pPR1058, pPR1287 which are subclones of E. coli O111 O antigen gene cluster. The thickened line is the region common to all clones. Broken lines show segments that are non-contiguous on the chromosome. The deduced restriction map for E. coli strain M92 is shown above.

Figure 2 shows a restriction mapping analysis of E. coli O111 O antigen gene cluster within the cosmid clone pPR1058. Restriction enzymes are: (B: *Bam*H1; Bg: *Bgl*II, E: *Eco*R1; H: *Hind*III; K: *Kpn*I; P: *Pst*I; S: *SalI* and X: *Xho*1. Plasmids pPR1230, pPR1231, and pPR1288 are deletion derivatives of pPR1058. Plasmids pPR 1237, pPR1238, pPR1239 and pPR1240 are in pUC19. Plasmids pPR1243, pPR1244, pPR1245, pPR1246 and pPR1248 are in pUC18, and pPR1292 is in pUC19. Plasmid pPR1270 is in pT7T319U. Probes 1, 2 and 3 were isolated as internal fragments of pPR1246, pPR1243 and pPR1237 respectively. Dotted lines indicate that subclone DNA extends to the left of the map into attached vector.

Figure 3 shows the structure of E. coli O111 O antigen gene cluster.

Figure 4 shows the structure of E. coli O157 O antigen gene cluster.

Figure 5 shows the structure S. enterica locus encoding the serogroup C2 O antigen gene cluster.

Figure 6 shows the structure S. enterica locus encoding the serogroup B O antigen gene cluster.

Figure 7 shows the nucleotide sequence of the E. coli O111 O antigen gene cluster. Note: (1) The first and last three bases of a gene are underlined and of italic respectively.; (2) The region which was previously sequenced by Bastin and Reeves 1995 "Sequence and anlysis of the O antigen gene (rfb) cluster of Eschericia coli o111" Gene 164: 17-23 is marked.

Figure 8 shows the nucleotide sequence of the E. coli O157 O antigen gene cluster. Note: (1) The first and last

three bases of a gene (region) are underlined and of *italic* respectively (2) The region previously sequenced by Bilge et al. 1996 "Role of the Eschericia coli O157-H7 O side chain in adherence and analysis of an rfb locus". Inf. and Immun 64:4795-4801 is marked.

Figure 9 shows the nucleotide sequence of S. enterica serogroup C2 O antigen gene cluster. Note: (1) The numbering is as in Brown et al. 1992. "Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B". Mol. Microbiol. 6: 1385-1394(2) The first and last three bases of a gene are underlined and in italics respectively.(3) Only that part of the group C2 gene cluster, which differs from that of group B, was sequenced and is presented here.

Figure 10 shows the nucleotide sequence of S. enterica serogroup B O antigen gene cluster Note: (1) The numbering is as in Jiang et al. 1991. "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2)". Mol. Microbiol. 5: 695-713. The first gene in the O antigen gene cluster is *rmlB* which starts at base 4099. (2) The first and last three bases of a gene are underlined and in italics respectively.

BEST METHOD FOR CARRYING OUT THE INVENTION

Materials and Methods-part 1

The experimental procedures for the isolation and characterisation of the E. coli O111 O antigen gene cluster (position 3,021-9,981) are according to Bastin D.A., *et al*. 1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol*. 5:9 2223-2231 and Bastin D.A. and Reeves, P.R. 1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111". *Gene* 164: 17-23.

A. Bacterial strains and growth media

Bacteria were grown in Luria broth supplemented as required.

B. Cosmids and phage

Cosmids in the host strain x2819 were repackaged *in vivo*. Cells were grown in 250mL flasks containing 30mL of culture, with moderate shaking at 30°C to an optical density of 0.3 at 580 nm. The defective lambda prophage was induced by heating in a water bath at 45°C for 15min followed by an incubation at 37°C with vigorous shaking for 2hr. Cells were then lysed by the addition of 0.3mL chloroform and shaking for a further 10min. Cell debris were removed from 1mL of lysate by a 5min spin in a microcentrifuge, and the supernatant removed to a fresh microfuge tube. One drop of chloroform was added then shaken vigorously through the tube contents.

C. DNA preparation

Chromosomal DNA was prepared from bacteria grown overnight at 37°C in a volume of 30mL of Luria broth. After harvesting by centrifugation, cells were washed and resuspended in 10mL of 50mMTris-HCl pH 8.0. EDTA was added and the mixture incubated for 20min. Then lysozyme was added and incubation continued for a further 10min. Proteinase K, SDS, and ribonuclease were then added and the mixture incubated for up to 2hr for lysis to occur. All incubations were at 37°C. The mixture was then heated to 65°C and extracted once with 8mL of phenol at the same temperature. The mixture was extracted once with 5mL of phenol/chloroform/iso-amyl alcohol at 4°C. Residual phenol was removed by two ether extractions. DNA was precipitated with 2 vols. of ethanol at 4°C, spooled and washed in 70% ethanol, resuspended in 1-2mL of TE and dialysed. Plasmid and cosmid DNA was prepared by a modification of the Birnboim and Doly method [Birnboim, H. C. And Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA *Nucl. Acid Res.* 7:1513-1523. The volume of culture was 10mL and the lysate was extracted with phenol/chloroform/iso-amyl alcohol before precipitation with isopropanol. Plasmid

DNA to be used as vector was isolated on a continuous caesium chloride gradient following alkaline lysis of cells grown in 1L of culture.

D. Enzymes and buffers.

Restriction endonucleases and DNA T4 ligase were purchased from Boehringer Mannheim (Castle Hill, NSW, Australia)or Pharmacia LKB (Melbourne, VIC Australia). Restriction enzymes were used in the recommended commercial buffer.

E. Construction of a gene bank.

Individual aliquots of M92 chromosomal DNA (strain Stoke W, from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen S, Denmark) were partially digested with 0.2U *Sau*3A1 for 1-15mins. Aliquots giving the greatest proportion of fragments in the size range of approximately 40-50kb were selected and ligated to vector pPR691 previously digested with *Bam*H1 and *Pvu*II. Ligation mixtures were packaged *in vitro* with packaging extract. The host strain for transduction was x2819 and recombinants were selected with kanamycin.

F. Serological procedures.

Colonies were screened for the presence of the O111 antigen by immunoblotting. Colonies were grown overnight, up to 100 per plate then transferred to nitrocellulose discs and lysed with 0.5N HCl. Tween 20 was added to TBS at 0.05% final concentration for blocking, incubating and washing steps. Primary antibody was <u>E. coli</u> O group 111 antiserum, diluted 1:800. The secondary antibody was goat anti-rabbit IgG labelled with horseradish peroxidase diluted 1:5000. The staining substrate was 4-chloro-1-napthol. Slide agglutination was performed according to the standard procedure.

G. Recombinant DNA methods.

Restriction mapping was based on a combination of standard methods including single and double digests and sub-cloning. Deletion derivatives of entire cosmids were produced as follows: aliquots of 1.8μg of cosmid DNA were

digested in a volume of 20µl with 0.25U of restriction enzyme for 5-80min. One half of each aliquot was used to check the degree of digestion on an agarose gel. The sample which appeared to give a representative range of fragments was ligated at 4°C overnight and transformed by the $CaCl_2$ method into JM109. Selected plasmids were transformed into sϕ174 by the same method. P4657 was transformed with pPR1244 by electroporation.

H. DNA hybridisation

Probe DNA was extracted from agarose gels by electroelution and was nick-translated using [α-32P]-dCTP. Chromosomal or plasmid DNA was electrophoresed in 0.8% agarose and transferred to a nitrocellulose membrane. The hybridisation and pre-hybridisation buffers contained either 30% or 50% formamide for low and high stringency probing respectively. Incubation temperatures were 42°C and 37°C for pre-hybridisation and hybridisation respectively. Low stringency washing of filters consisted of 3 x 20min washes in 2 x SSC and 0.1% SDS. High-stringency washing consisted of 3 x 5min washes in 2 x SSC and 0.1% SDS at room temperature, a 1hr wash in 1 x SSC and 0.1% SDS at 58°C and 15min wash in 0.1 x SSC and 0.1% SDS at 58°C.

I. Nucleotide sequencing of E. coli O111 O antigen gene cluster (position 3,021-9,981)

Nucleotide sequencing was performed using an ABI 373 automated sequencer (CA, USA). The region between map positions 3.30 and 7.90 was sequenced using uni-directional exonuclease III digestion of deletion families made in PT7T3190 from clones pPR1270 and pPR1272. Gaps were filled largely by cloning of selected fragments into M13mp18 or M13mp19. The region from map positions 7.90-10.2 was sequenced from restriction fragments in M13mp18 or M13mp19. Remaining gaps in both the regions were filled by priming from synthetic oligonucleotides complementary to determined positions along the sequence,

using a single stranded DNA template in M13 or phagemid. The oligonucleotides were designed after analysing the adjacent sequence. All sequencing was performed by the chain termination method. Sequences were aligned using SAP [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing". *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231]. The program NIP [Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961] was used to find open reading frames and translate them into proteins.

J. Isolation of clones carrying E. coli 0111 O antigen gene cluster

The E. coli O antigen gene cluster was isolated according to the method of Bastin D.A., *et al.* [1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli 0111 strain". *Mol. Microbiol.* 5(9), 2223-2231]. Cosmid gene banks of M92 chromosomal DNA were established in the *in vivo* packaging strain x2819. From the genomic bank, 3.3 x 10^3 colonies were screened with E. coli 0111 antiserum using an immuno-blotting procedure: 5 colonies (pPR1054, pPR1055, pPR1056, pPR1058 and pPR1287) were positive. The cosmids from these strains were packaged *in vivo* into lambda particles and transduced into the E. coli deletion mutant Sϕ174 which lacks all O antigen genes. In this host strain, all plasmids gave positive agglutination with 0111 antiserum. An *Eco* R1 restriction map of the 5 independent cosmids showed that they have a region of approximately 11.5 kb in common (Figure 1). Cosmid pPR1058 included sufficient flanking DNA to identify several chromosomal markers linked to O antigen gene cluster and was selected for analysis of the O antigen gene cluster region.

K. Restriction mapping of cosmid pPR1058

Cosmid pPR1058 was mapped in two stages. A preliminary map was constructed first, and then the region between map positions 0.00 and 23.10 was mapped in detail, since it was shown to be sufficient for O111 antigen expression. Restriction sites for both stages are shown in Figure 2. The region common to the five cosmid clones was between map positions 1.35 and 12.95 of pPR1058.

To locate the O antigen gene cluster within pPR1058, pPR1058 cosmid was probed with DNA probes covering O antigen gene cluster flanking regions from S. enterica LT2 and E. coli K-12. Capsular polysaccharide (*cps*) genes lie upstream of O antigen gene cluster while the gluconate dehydrogenase (*gnd*) gene and the histidine (*his*) operon are downstream, the latter being further from the O antigen gene cluster. The probes used were pPR472 (3.35kb), carrying the *gnd* gene of LT2, pPR685 (5.3kb) carrying two genes of the *cps* cluster, *cpsB* and *cpsG* of LT2, and K350 (16.5kb) carrying all of the *his* operon of K-12. Probes hybridised as follows: pPR472 hybridised to 1.55kb and 3.5 kb (including 2.7 kb of vector) fragments of *Pst*1 and *Hind*III double digests of pPR1246 (a *Hind*III/*Eco*R1 subclone derived from pPR1058, Figure 2), which could be located at map positions 12.95-15.1; pPR685 hybridised to a 4.4 kb *Eco*R1 fragment of pPR1058 (including 1.3 kb of vector) located at map position 0.00-3.05; and K350 hybridised with a 32kb *Eco*R1 fragment of pPR1058 (including 4.0kb of vector), located at map position 17.30-45.90. Subclones containing the presumed *gnd* region complemented a $gnd^{-}edd^{-}$ strain GB23152. On gluconate bromothymol blue plates, pPR1244 and pPR1292 in this host strain gave the green colonies expected of a $gnd^{-}edd^{-}$ genotype. The his^{+} phenotype was restored by plasmid pPR1058 in the *his* deletion strain Sϕ174 on minimal medium plates, showing that the plasmid carries the entire his operon.

It is likely that the O antigen gene cluster region lies between *gnd* and *cps*, as in other E. coli and S. enterica strains, and hence between the approximate map

positions 3.05 and 12.95. To confirm this, deletion derivatives of pPR1058 were made as follows: first, pPR1058 was partially digested with *Hind*III and self ligated. Transformants were selected for kanamycin resistance and screened for expression of O111 antigen. Two colonies gave a positive reaction. *Eco*R1 digestion showed that the two colonies hosted identical plasmids, one of which was designated pPR1230, with an insert which extended from map positions 0.00 to 23.10. Second pPR1058 was digested with *Sal*1 and partially digested with *Xho*1 and the compatible ends were re-ligated. Transformants were selected with kanamycin and screened for O111 antigen expression. Plasmid DNA of 8 positively reacting clones was checked using *Eco*R1 and *Xho*1 digestion and appeared to be identical. The cosmid of one was designated pPR1231. The insert of pPR1231 contained the DNA region between map positions 0.00 and 15.10. Third, pPR1231 was partially digested with *Xho*1, self-ligated, and transformants selected on spectinomycin/ streptomycin plates. Clones were screened for kanamycin sensitivity and of 10 selected, all had the DNA region from the *Xho*1 site in the vector to the *Xho*1 site at position 4.00 deleted. These clones did not express the O111 antigen, showing that the *Xho*1 site at position 4.00 is within the O antigen gene cluster. One clone was selected and named pPR1288. Plasmids pPR1230, pPR1231, and pPR1288 are shown in Figure 2.

L. Analysis of the E. coli O111 O antigen gene cluster (position 3,021-9,981) nucleotide sequence data

Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111". *Gene* 164: 17-23] partially characterised the E. coli O111 O antigen gene cluster by sequencing a fragment from map position 3,021-9,981. Figure 3 shows the gene organisation of position 3,021-9,981 of E. coli O111 O antigen gene cluster. *orf3* and *orf6* have high level amino acid identity with *wcaH* and *wcaG* (46.3% and 37.2% respectively), and are likely to be similar in function to

sugar biosynthetic pathway genes in the E. coli K-12 colanic gene cluster. *orf4* and *orf5* show high levels of amino acid homology to *manC* and *manB* genes respectively. *orf7* shows high level homology with *rfbH* which is an abequose pathway gene. *orf8* encodes a protein with 12 transmembrane segments and has similarity in secondary structure to other *wzx* genes and is likely therefore to be the O antigen flippase gene.

Materials and Methods-part 2

A. Nucleotide sequencing of 1 to 3,020 and 9,982 to 14,516 of the E. coli O111 O antigen gene cluster

The sub clones which contained novel nucleotide sequences, pPR1231 (map position 0 and 1,510), pPR1237 (map position -300 to 2,744), pPR1239 (map position 2,744 to 4,168), pPR1245 (map position 9,736 to 12,007) and pPR1246 (map position 12,007 to 15,300) (Figure 2), were characterised as follows: the distal ends of the inserts of pPR1237, pPR1239 and pPR1245 were sequenced using the M13 forward and reverse primers located in the vector. PCR walking was carried out to sequence further into each insert using primers based on the sequence data and the primers were tagged with M13 forward or reverse primer sequences for sequencing. This PCR walking procedure was repeated until the entire insert was sequenced. pPR1246 was characterised from position 12,007 to 14,516. The DNA of these sub clones was sequenced in both directions. The sequencing reactions were performed using the dideoxy termination method and thermocycling and reaction products were analysed using fluorescent dye and an ABI automated sequencer (CA, USA).

B. Analysis of the E. coli O111 O antigen gene cluster (positions 1 to 3,020 and 9,982 to 14,516 of SEQ ID NO:1) nucleotide sequence data

The gene organisation of regions of E. coli O111 O antigen gene cluster which were not characterised by Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111." *Gene*

164: 17-23],(positions 1 to 3,020 and 9,982 to 14,516) is shown in Figure 3. There are two open reading frames in region 1. Four open reading frames are predicted in region 2. The position of each gene is listed in Table 5.

The deduced amino acid sequence of *orf1* (*wbdH*) shares about 64% similarity with that of the *rfp* gene of Shigella dysenteriae. Rfp and WbdH have very similar hydrophobicity plots and both have a very convincing predicted transmembrane segment in a corresponding position. *rfp* is a galactosyl transferase involved in the synthesis of LPS core, thus *wbdH* is likely to be a galactosyl transferase gene. *orf2* has 85.7% identity at amino acid level to the *gmd* gene identified in the E. coli K-12 colanic acid gene cluster and is likely to be a *gmd* gene. *orf9* encodes a protein with 10 predicted transmembrane segments and a large cytoplasmic loop. This inner membrane topology is a characteristic feature of all known O antigen polymerases thus it is likely that *orf9* encodes an O antigen polymerase gene, *wzy*. *orf10* (*wbdL*) has a deduced amino acid sequence with low homology with *Lsi2* of *Neisseria gonorrhoeae*. *Lsi2* is responsible for adding GlcNAc to galactose in the synthesis of lipooligosaccharide. Thus it is likely that *wbdL* is either a colitose or glucose transferase gene. *orf11* (*wbdM*) shares high level nucleotide and amino acid similarity with TrsE of *Yersinia enterocholitica*. TrsE is a putative sugar transferase thus it is likely that *wbdM* encodes the colitose or glucose transferase.

In summary three putative transferase genes and an O antigen polymerase gene were identified at map position 1 to 3,020 and 9,982 to 14,516 of E. coli O111 O antigen gene cluster. A search of GenBank has shown that there are no genes with significant similarity at the nucleotide sequence level for two of the three putative transferase genes or the polymerase gene. SEQ ID NO:1 and Figure 7 provide the nucleotide sequence of the O111 antigen gene cluster.

Materials and Methods-part 3

A. PCR amplification of O157 antigen gene cluster from an E. coli O157:H7 strain (Strain C664-1992, from Statens Serum Institut, 5 Artillerivej, 2300, Copenhagen S, Denmark)

E. coli O157 O antigen gene cluster was amplified by using long PCR [Cheng et al. 1994, Effective amplification of long targets from cloned inserts and human and genomic DNA" P.N.A.S. USA 91: 5695-569] with one primer (primer #412: att ggt agc tgt aag cca agg gcg gta gcg t) based on the JumpStart sequence usually found in the promoter region of O antigen gene clusters [Hobbs, *et al.* 1994 "The JumpStart sequence: a 39 bp element common to several polysaccharide gene clusted" Mol. Microbiol. 12: 855-856], and another primer #482 (cac tgc cat acc gac gac gcc gat ctg ttg ctt gg) based on the *gnd* gene usually found downstream of the O antigen gene cluster. Long PCR was carried out using the Expand Long Template PCR System from Boehringer Mannheim (Castle Hill NSW Australia), and products, 14 kb in length, from several reactions were combined and purified using the Promega Wizard PCR preps DNA purification System (Madison WI USA). The PCR product was then extracted with phenol and twice with ether, precipitated with 70% ethanol, and resuspended in 40μL of water.

B. Construction of a random DNase I bank:

Two aliquots containing about 150ng of DNA each were subjected to DNase I digestion using the Novagen DNase I Shotgun Cleavage (Madison WI USA) with a modified protocol as described. Each aliquot was diluted into 45μl of 0.05M Tris -HCl (pH7.5), 0.05mg/mL BSA and 10mM $MnCl_2$. 5μL of 1:3000 or 1:4500 dilution of DNaseI (Novagen) (Madison WI USA) in the same buffer was added into each tube respectively and 10μl of stop buffer (100mM EDTA), 30% glycerol, 0.5% Orange G, 0.075% xylene and cyanol (Novagen) (Madison WI USA) was added after incubation at 15°C for 5 min. The DNA from the two DNaseI reaction

tubes were then combined and fractionated on a 0.8% LMT agarose gel, and the gel segment with DNA of about 1kb in size (about 1.5mL agarose) was excised. DNA was extracted from agarose using Promega Wizard PCR Preps DNA Purification (Madison WI USA) and resuspended in 200 µL water, before being extracted with phenol and twice with ether, and precipitated. The DNA was then resuspended in 17.25 µL water and subjected to T4 DNA polymerase repair and single dA tailing using the Novagen Single dA Tailing Kit (Madison WI USA). The reaction product (85µl containing about 8ng DNA) was then extracted with chloroform:isoamyl alcohol (24:1) once and ligated to 3x 10^{-3} pmol pGEM-T (Promega) (Madison WI USA) in a total volume of 100µL. Ligation was carried out overnight at 4°C and the ligated DNA was precipitated and resuspended in 20µL water before being electroporated into <u>E. coli</u> strain JM109 and plated out on BCIG-IPTG plates to give a bank.

C. Sequencing

DNA templates from clones of the bank were prepared for sequencing using the 96-well format plasmid DNA miniprep kit from Advanced Genetic Technologies Corp (Gaithersburg MD USA) The inserts of these clones were sequenced from one or both ends using the standard M13 sequencing primer sites located in the pGEM-T vector. Sequencing was carried out on an ABI377 automated sequencer (CA USA) as described above, after carrying out the sequencing reaction on an ABI Catalyst (CA USA). Sequence gaps and areas of inadequate coverage were PCR amplified directly from O157 chromosomal DNA using primers based on the already obtained sequencing data and sequenced using the standard M13 sequencing primer sites attached to the PCR primers.

D. Analysis of the <u>E. coli</u> O157 O antigen gene cluster nucleotide sequence data

Sequence data were processed and analysed using the

Staden programs [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231; Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961]. Figure 4 shows the structure of E. coli O157 O antigen gene cluster. Twelve open reading frames were predicted from the sequence data, and the nucleotide and amino acid sequences of all these genes were then used to search the GenBank database for indication of possible function and specificity of these genes. The position of each gene is listed in Table 6. The nucleotide sequence is presented in SEQ ID NO:2 and Figure 8.

orfs 10 and 11 showed high level identity to *manC* and *manB* and were named *manC* and *manB respectively. orf7* showed 89% identity (at amino acid level) to the *gmd* gene of the E. coli colanic acid capsule gene cluster (Stevenson G., K. *et al.* 1996 "Organisation of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid".J. Bacteriol. 178:4885-4893) and was named *gmd*. *orf8* showed 79% and 69% identity (at amino acid level) respectively to *wcaG* of the E. coli colanic acid capsule gene cluster and to *wbcJ* (*orf14.8*) gene of the Yersinia enterocolitica O8 O antigen gene cluster (Zhang, L. *et al.* 1997 "Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of Y. enterocolitica serotype O8".Mol. Microbiol. 23:63-76). Colanic acid and the Yersinia O8 O antigen both contain fucose as does the O157 O antigen. There are two enzymatic steps required for GDP-L-fucose synthesis from GDP-4-keto-6-deoxy-D-mannose, the product of the *gmd* gene product. However, it has been shown recently (Tonetti, M *et al.* 1996 Synthesis of GDP-L-fucose by the human FX protein J. Biol. Chem. 271:27274-27279) that the human FX

protein has "significant homology" with the *wcaG* gene (referred to as *Yefb* in that paper), and that the FX protein carries out both reactions to convert GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose. We believe that this makes a very strong case for *orf8* carrying out these two steps and propose to name the gene *fcl*. In support of the one enzyme carrying out both functions is the observation that there are no genes other than *manB*, *manC*, *gmd* and *fcl* with similar levels of similarity between the three bacterial gene clusters for fucose containing structures.

orf5 is very similar to *wbeE* (*rfbE*) of Vibrio cholerae 01, which is thought to be the perosamine synthetase, which converts GDP-4-keto-6-deoxy-D-mannose to GDP-perosamine (Stroeher, U.H *et al.* 1995 "A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of Vibrio cholerae" 01. Gene 166: 33-42). V. cholerae 01 and E. coli 0157 O antigens contain perosamine and N-acetyl-perosamine respectively. The V. cholerae 01 *manA*, *manB*, *gmd* and *wbeE* genes are the only genes of the V. cholerae 01 gene cluster with significant similarity to genes of the E. coli 0157 gene cluster and we believe that our observations both confirm the prediction made for the function of *wbe* of V. cholerae, and show that *orf5* of the 0157 gene cluster encodes GDP-perosamine synthetase. *orf5* is therefore named *per*. *orf5* plus about 100bp of the upstream region (postion 4022-5308)was previously sequenced by Bilge, S.S. *et al.* [1996 "Role of the Escherichia coli 0157-H7 O side chain in adherence and analysis of an *rfb* locus".Infect. Immun. 64:4795-4801].

orf12 shows high level similarity to the conserved region of about 50 amino acids of various members of an acetyltransferase family (Lin, W., *et al.* 1994 "Sequence analysis and molecular characterisation of genes required for the biosynthesis of type 1 capsular polysaccharide in Staphylococcus aureus". J. Bateriol. 176: 7005-7016) and we believe it is the N-acetyltransferase to convert GDP-perosamine to GDP-perNAc. *orf12* has been named *wbdR*.

The genes *manB*, *manC*, *gmd*, *fcl*, *per* and *wbdR* account for all of the expected biosynthetic pathway genes of the O157 gene cluster.

The remaining biosynthetic step(s) required are for synthesis of UDP-GalNAc from UDP-Glc. It has been proposed (Zhang, L., *et al.* 1997 "Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of Yersinia enterocolitica serotype O8".Mol. Microbiol. 23:63-76) that in Yersinia enterocolitica UDP-GalNAc is synthesised from UDP-GlcNAc by a homologue of galactose epimerase (GalE), for which there is a *galE* like gene in the Yersinia enterocolitica O8 gene cluster. In the case of O157 there is no *galE* homologue in the gene cluster and it is not clear how UDP-GalNAc is synthesised. It is possible that the galactose epimerase encoded by the *galE* gene in the *gal* operon, can carry out conversion of UDP-GlcNAc to UDP-GalNAc in addition to conversion of UDP-Glc to UDP-Gal. There do not appear to be any gene(s) responsible for UDP-GalNAc synthesis in the O157 gene cluster.

orf4 shows similarity to many *wzx* genes and is named *wzx* and *orf2* which shows similarity of secondary structure in the predicted protein to other *wzy* genes and is for that reason named *wzy*.

The *orf1*, *orf3* and *orf6* gene products all have characteristics of transferases, and have been named *wbdN*, *wbdO* and *wbdP* respectively. The O157 O antigen has 4 sugars and 4 transferases are expected. The first transferase to act would put a sugar phosphate onto undecaprenol phosphate. The two transferases known to perform this function, WbaP (RfbP) and WecA (Rfe) transfer galactose phosphate and N-acetyl-glucosamine phosphate respectively to undecaprenol phosphate. Neither of these sugars is present in the O157 structure.

Further, none of the presumptive transferases in the O157 gene cluster has the transmembrane segments found in *WecA* and WbaP which transfer a sugar phosphate to undecaprenol phosphate and expected for any protein which

transferred a sugar to undecaprenol phosphate which is embedded within the membrane.

The *WecA* gene which transfers GlcNAc-P to undecaprenol phosphate is located in the Enterobactereal Common Antigen (ECA) gene cluster and it functions in ECA synthesis in most and perhaps all E. coli strains, and also in O antigen synthesis for those strains which have GlcNAc as the first sugar in the O unit.

It appears that WecA acts as the transferase for addition of GalNAc-1-P to undecaprenol phosphate for the Yersinia enterocolitica O8 O antigen [Zhang et al.1997 "Molecular and chemical characterisation of the lipopolysaccharide O antigen and its role in the virulence of Yersinia enterocolitica serotype O8" Mol. Microbiol. 23: 63-76.] and perhaps does so here as the O157 structure includes GalNAc. *WecA* has also been reported to add Glucose-1-P phosphate to undecaprenol phosphate in E. coli O8 and O9 strains, and an alternative possibility for transfer of the first sugar to undecaprenol phosphate is *WecA* mediated transfer of glucose, as there is a glucose residue in the O157 O antigen. In either case the requisite number of transferase genes are present if GalNAc or Glc is transferred by *WecA* and the side chain Glc is transferred by a transferase outside of the O antigen gene cluster.

orf9 shows high level similarity (44% identity at amino acid level, same length) with *wcaH* gene of the E. coli colanic acid capsule gene cluster. The function of this gene is unknown, and we give *orf9* the name *wbdQ*.

The DNA between *manB* and *wdbR* has strong sequence similarity to one of the H-repeat units of E. coli K12. Both of the inverted repeat sequences flanking this region are still recognisable, each with two of the 11 bases being changed. The H-repeat associated protein encoding gene located within this region has a 267 base deletion and mutations in various positions. It seems that the H-repeat unit has been associated with this gene cluster for a long period of time since it translocated to the gene

cluster, perhaps playing a role in assembly of the gene cluster as has been proposed in other cases.

Materials and Methods - part 4

To test our hypothesis that O antigen genes for transferases and the *wzx*, *wzy* genes were more specific than pathway genes for diagnostic PCR, we first carried out PCR using primers for all the E. coli 016 O antigen genes (Table 4). The PCR was then carried out using PCR primers for E. coli 0111 transferase, *wzx* and *wzy* genes (Table 5, 5A). PCR was also carried out using PCR primers for the E. coli 0157 transferase, *wzx* and *wzy* genes (Table 6, 6A).

Chromosomal DNA from the 166 serotypes of E. coli available from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen Denmark was isolated using the Promega Genomic (Madison WI USA) isolation kit. Note that 164 of the serogroups are described by Ewing W. H.: Edwards and Ewings "Identification of the Enterobacteriacea" Elsevier, Amsterdam 1986 and that they are numbered 1-171 with numbers 31, 47, 67, 72, 93, 94 and 122 no longer valid. Of the two serogroup 19 strains we used 19ab strain F8188-41. Lior H. 1994 ["Classification of Eschericia coli In Eschericia coli in domestic animals and humans pp 31-72. Edited by C.L. Gyles CAB international] adds two more numbered 172 and 173 to give the 166 serogroups used. Pools containing 5 to 8 samples of DNA per pool were made. Pool numbers 1 to 19 (Table 1) were used in the E. coli 0111 and 0157 assay. Pool numbers 20 to 28 were also used in the 0111 assay, and pool numbers 22 to 24 contained E. coli 0111 DNA and were used as positive controls (Table 2). Pool numbers 29 to 42 were also used in the 0157 assay, and pool numbers 31 to 36 contained E. coli 0157 DNA, and were used as positive controls (Table 3). Pool numbers 2 to 20, 30, 43 and 44 were used in the E. coli 016 assay (Tables 1 to 3). Pool number 44 contained DNA of E. coli K-12 strains C600 and WG1 and was used as a positive control as between them they have all of the E.

coli K-12 O16 O antigen genes.

PCR reactions were carried out under the following conditions: denaturing 94°C/30″; annealing, temperature varies (refer to Tables 4 to 8)/30″; extension, 72°C/1′; 30 cycles. PCR reaction was carried out in an volume of 25μL for each pool. After the PCR reaction, 10μL PCR product from each pool was run on an agarose gel to check for amplified DNA.

Each E. coli and S. enterica chromosomal DNA sample was checked by gel electrophoresis for the presence of chromosomal DNA and by PCR amplification of the E. coli or S. enterica *mdh* gene using oligonucleotides based on E. coli K-12 or Salmonella enterica LT2 [Boyd *et al.* (1994) "Molecular genetic basis of allelic polymorphism in malate degydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*" Proc. Nat. Acad. Sci. USA. 91:1280-1284.] Chromosomal DNA samples from other bacteria were only checked by gel electrophoresis of chromosomal DNA.

A. Primers based on E. coli O16 O antigen gene cluster sequence.

The O antigen gene cluster of E. coli O16 was the only typical E. coli O antigen gene cluster that had been fully sequenced prior to that of O111, and we chose it for testing our hypothesis. One pair of primers for each gene was tested against pools 2 to 20, 30 and 43 of E. coli chromosomal DNA. The primers, annealing temperatures and functional information for each gene are listed in Table 4.

For the five pathway genes, there were 17/21, 13/21, 0/21, 0/21, 0/21 positive pools for *rmlB*, *rmlD*, *rmlA*, *rmlC* and *glf* respectively (Table 4). For the *wzx*, wzy and three transferase genes there were no positives amongst the 21 pools of E. coli chromosomal DNA tested (Table 4). In each case the #44 pool gave a positive result.

B. Primers based on the E. coli 0111 O antigen gene clsuter sequence.

One to four pairs of primers for each of the transferase, *wzx* and *wzy* genes of 0111 were tested against the pools 1 to 21 of E. coli chromosomal DNA (Table 5). For *wbdH*, four pairs of primers, which bind to various regions of this gene, were tested and found to be specific for 0111 as there was no amplified DNA of the correct size in any of those 21 pools of E. coli chromosomal DNA tested. Three pairs of primers for *wbdM* were tested, and they are all specific although primers #985/#986 produced a band of the wrong size from one pool. Three pairs of primers for *wzx* were tested and they all were specific. Two pairs of primers were tested for *wzy*, both are specific although #980/#983 gave a band of the wrong size in all pools. One pair of primers for *wbdL* was tested and found unspecific and therefore no further test was carried out. Thus, *wzx*, *wzy* and two of the three transferase genes are highly specific to 0111. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and positions for each gene are in (Table 5).

The 0111 assay was also performed using pools including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii and Salmonella enterica strains (Table 5A). None of the oligonucleotides derived from *wbdH*, *wzx*, *wzy* or *wbdM* gave amplified DNA of the correct size with these pools. Notably, pool number 25 includes S. enterica Adelaide which has the same O antigen as E. coli 0111: this pool did not give a positive PCR result for any primers tested indicating that these genes are highly specific for E. coli 0111.

Each of the 12 pairs binding to *wbdH*, *wzx*, *wzy* and *wbdM* produces a band of predicted size with the pools containing 0111 DNA (pools number 22 to 24). As pools 22 to 24 included DNA from all strains present in pool 21 plus 0111 strain DNA (Table 2), we conclude that the 12

pairs of primers all give a positive PCR test with each of three unrelated O111 strains but not with any other strains tested. Thus these genes are highly specific for E. coli O111.

C. Primers based on the E. coli O157 O antigen gene cluster sequence.

Two or three primer pairs for each of the transferase, *wzx* and *wzy* genes of O157 were tested against E. coli chromosomal DNA of pools 1 to 19, 29 and 30 (Table 6). For *wbdN*, three pairs of primers, which bind to various regions of this gene, were tested and found to be specific for O157 as there was no amplified DNA in any of those 21 pools of E. coli chromosomal DNA tested. Three pairs of primers for *wbdO* were tested, and they are all specific although primers # 1211/#1212 produced two or three bands of the wrong size from all pools. Three pairs of primers were tested for *wbdP* and they all were specific. Two pairs of primers were tested for *wbdR* and they were all specific. For *wzy*, three pairs of primers were tested and all were specific although primer pair #1203/#1204 produced one or three bands of the wrong size in each pool. For *wzx*, two pairs of primers were tested and both were specific although primer pair #1217/#1218 produced 2 bands of wrong size in 2 pools, and 1 band of wrong size in 7 pools. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and function information for each gene are in Table 6.

The O157 assay was also performed using pools 37 to 42, including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii, Yersinia enterocolitica O9, Brucella abortus and Salmonella enterica strains (Table 6A). None of the oligonucleotides derived from *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* or *wbdR* reacted specifically with these pools, except that primer pair #1203/#1204 produced two bands with Y. enterocolitica O9

and one of the bands is of the same size with that from the positive control. Primer pair #1203/#1204 binds to *wzy*. The predicted secondary structures of Wzy proteins are generally similar, although there is very low similarity at amino acid or DNA level among the sequenced *wzy* genes. Thus, it is possible that Y. enterolcolitica O9 has a *wzy* gene closely related to that of E. coli O157. It is also possible that this band is due to chance hybridization of another gene, as the other two *wzy* primer pairs (#1205/#1206 and #1207/#1208) did not produce any band with Y. enterocolitica O9. Notably, pool number 37 includes S. enterica Landau which has the same O antigen as E. coli O157, and pool 38 and 39 contain DNA of B. abortus and Y. enterocolitica O9 which cross react serologically with E. coli O157. This result indicates that these genes are highly O157 specific, although one primer pair may have cross reacted with Y. enterocolitica O9.

Each of the 16 pairs binding to *wbdN, wzx, wzy, wbdO, wbdP* and *wbdR* produces a band of predicted size with the pools containing O157 DNA (pools number 31 to 36). As pool 29 included DNA from all strains present in pools 31 to 36 other than O157 strain DNA (Table 3), we conclude that the 16 pairs of primers all give a positive PCR test with each of the five unrelated O157 strains.

Thus PCR using primers based on genes *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* and *wbdR* is highly specific for E. coli O157, giving positive results with each of six unrelated O157 strains while only one primer pair gave a band of the expected size with one of three strains with O antigens known to cross-react serologically with E. coli O157.

D. Primers based on the Salmonella enterica serotype C2 and B O antigen gene cluster sequences.

We also performed a PCR using primers for the S. enterica C2 and B serogroup transferases, *wzx*, *wzy* and genes (Tables 7 to 9). The nucleotide sequences of C2

and B O antigen gene clusters are listed as SEQ ID NO: 3 (Fig. 9) and SEQ ID NO:4 (Fig. 10)respectively. Chromosomal DNA from all the 46 serotypes of Salmonella enterica (Table 9) was isolated using the Promega Genomic isolation kit, 7 pools of 4 to 8 samples per pool were made. Salmonella enterica serotype B or C2 DNA was omitted from the pool for testing primers of 46 respective serotypes but added to a pool containing 6 other samples to give pool number 8 for use as a positive control.

PCR reactions were carried out under the following conditions: denaturing, 94°C/30″; annealing, temperature varies (see below)/30″; extension, 72°C/1′; 30 cycles. PCR reaction was carried out in a volume of 25μL for each pool. After the PCR reaction, 10μL PCR product from each pool was run on an agarose gel to check for amplified DNA. For pools which gave a band of correct size, PCR was repeated using individual chromosomal samples of that pool, and agarose gel was run to check for amplified DNA from each sample.

The Salmonella enterica serotype B O antigen gene cluster (of strain LT2) was the first O antigen gene cluster to be fully sequenced, and the function of each gene has been identified experimentally [Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K., and Reeves, P. R. (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2)." Mol. Microbiol. **5**(3), 695-713; Liu, D., Cole, R., and Reeves, P. R. (1996). "An O antigen processing function for Wzx(RfbX): a promising candidate for O-unit flippase" J. Bacteriol., **178**(7),2102-2107; Liu, D., Haase, A. M., Lindqvist, L., Lindberg, A. A., and Reeves, P. R. (1993). "Glycosyl transferases of O-antigen biosynthesis in *S. enterica* : identification and characterisation of transferase genes of groups B, C2 and E1." J. Bacteriol., **175**, 3408-3413; Liu, D., Lindquist, L., and Reeves P. R. (1995). "Transferases of O-antigen biosynthesis in *Salmonella enterica*: dideoxhexosyl

transferases of groups B and C2 and acetyltransferase of group C2." J. Bacteriol., **177**, 4084-4088; Romana, L. K., Santiago, F. S., and Reeves, P. R. (1991). "High level expression and purification dThymidine-diphospho-D-glucose 4,6 dehydratase (*rfbB*) from *Salmonella* serovar *typhimurium* LT2." BBRC, **174**, 846-852]. One pair of primers for each of the pathway genes and *wbaP* was tested against the pools of Salmonella enterica DNA, two to three pairs of primers for each of the other transferases and *wzx* genes were also tested. See Table 8 for a list of primers and functional information of each gene, as well as the annealing temperature of the PCR reaction for each pair of primers.

For pathway genes of group B strain LT2, there are 19/45, 14/45, 15/45, 12/45, 6/45, 6/45, 6/45, 6/45, 1/45, 9/45, 8/45 positives for *rmlB*, *rmlD*, *rmlA*, *rmlC*, *ddhD*, *ddhA*, *ddhB*, *ddhC*, *abe*, *manC*, and *manB* repsectively (Table 9).

For the LT2 *wzx* gene we used three primer pairs each of which gave 1/45 positive. For the 4 transferase genes we used a total of 9 primer pairs. 2 primer pairs for *wbaV* gave 2/90 positives. For 3 primer pairs of *wbaN*, 11/135 gave a positive result. For the *wbaP* primer pair 10/45 gave a positive result (Table 9).

The experimental data show that oligonucleotides derived from the *wzx* and *wbaV* group B O antigen genes are specific for group B O antigen amongst all 45 Salmonella enterica O antigen groups except O group 67. The oligonucleotides derived from Salmonella enterica B group *wbaN* and *wbaU* genes detected B group O antigen and also produced positive results with groups A, D1 and D3. *WbaU* encodes a transferase for a Mannose α(1-4) Mannose linkage and is expressed in groups A, B and D1 while *wbaN*, which encodes a transferase for Rhamnose α(1-3) Galactose linkage is present in groups A, B, D1, D2, D3 and E1. This accounts for the positive results with the group B *wbaU* and *wbaN* genes. The *wbaN* gene of groups E and D2 has considerable sequence differences from that of groups A,

B, D1 and D3 and this accounts for the positive results only with groups B, D1 and D3.

The Salmonella enterica B primers derived from *wzx* and transferase genes produced a positive result with Salmonella enterica 067. We find that Salmonella enterica 067 has all the genes of the group B O antigen cluster. There are several possible explanations for this finding including the possibility that the gene cluster is not functional due to mutation and the group 067 antigenicity is due to another antigen, or the O antigen is modified after synthesis such that its antigenicity is changed. Salmonella enterica 067 would therefore be scored as Salmonella enterica group B in the PCR diagnostic assay. However, this is of little importance because Salmonella enterica 067 is a rare O antigen and only one (serovar Crossness) of the 2324 known serovars has the 067 serotype [Popoff M.Y. *et al* (1992) "Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France], and serovar Crossness had only been isolated once [M. Popoff, personal communication].

The Salmonella enterica B primers derived from *wbaP* reacted with group A, C2, D1, D2, D3, E1, 54, 55, 67 and E4 O antigen groups. *WbaP* encodes the galactosyl transferase which initiates O unit synthesis by transfer of Galactose phosphate to the lipid carrier Undecaprenol phosphate. This reaction is common to the synthesis of several O antigens. As such *wbaP* is distinguished from other transferases of the invention as it does not make a linkage within an O antigen.

We also tested 20 primer pairs for the *wzx*, *wzy* and 5 transferase genes of serotype C2 and found no positives in all the 7 pools (Table 7).

Groups A, B, D1, D2, D3, C2 and E1 share many genes in common. Some of these genes occur with more than one sequence in which case each specific sequence can be named after one of the serogroups in which it occurs. The

distribution of these sequence specificities is shown in Table 10. The inventors have aligned the nucleotide sequences of Salmonella enterica *wzy*, *wzx* genes and transferase genes so as to determine specific combinations of nucleic acid molecules which can be employed to specifically detect and identify the Salmonella enterica groups A, B, D1, D2, D3, C2 and E1 (Table 10). The results show that many of the O antigen groups can be detected and identified using a single specific nucleic acid molecule although other groups in particular D2 and E1, and A and D1 require a panel of nucleic acid molecules derived from a combination of genes.

It will be understood that in carrying out the methods of the invention with respect to the testing of particular sample types including samples from food, patients and faeces the samples are prepared by routine techniques routinely used in the preparation of such samples for DNA based testing.

TABLE 1

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
1	*E. coli* type strains for O serotypes 1, 2, 3, 4, 10, 16, 18 and 39	IMVS[a]
2	*E. coli* type strains for O serotypes 40, 41, 48, 49, 71, 73, 88 and 100	IMVS
3	*E. coli* type strains for O serotypes 102, 109, 119, 120, 121, 125, 126 and 137	IMVS
4	*E. coli* type strains for O serotypes 138, 139, 149, 7, 5, 6, 11 and 12	IMVS
5	*E. coli* type strains for O serotypes 13, 14, 15, 17, 19ab, 20, 21 and 22	IMVS
6	*E. coli* type strains for O serotypes 23, 24, 25, 26, 27, 28, 29 and 30	IMVS
7	*E. coli* type strains for O serotypes 32, 33, 34, 35, 36, 37, 38 and 42	IMVS
8	*E. coli* type strains for O serotypes 43, 44, 45, 46, 50, 51, 52 and 53	IMVS
9	*E. coli* type strains for O serotypes 54, 55, 56, 57, 58, 59, 60 and 61	IMVS
10	*E. coli* type strains for O serotypes 62, 63, 64, 65, 66, 68, 69 and 70	IMVS
11	*E. coli* type strains for O serotypes 74, 75, 76, 77, 78, 79, 80 and 81	IMVS
12	*E. coli* type strains for O serotypes 82, 83, 84, 85, 86, 87, 89 and 90	IMVS
13	*E. coli* type strains for O serotypes 91, 92, 95, 96, 97, 98, 99 and 101	IMVS
14	*E. coli* type strains for O serotypes 103, 104, 105, 106, 107, 108 and 110	IMVS
15	*E. coli* type strains for O serotypes 112, 162, 113, 114, 115, 116, 117 and 118	IMVS
16	*E. coli* type strains for O serotypes 123, 165, 166, 167, 168, 169, 170 and 171	See b
17	*E. coli* type strains for O serotypes 172, 173, 127, 128, 129, 130, 131 and 132	See c
18	*E. coli* type strains for O serotypes 133, 134, 135, 136, 140, 141, 142 and 143	IMVS
19	*E. coli* type strains for O serotypes 144, 145, 146, 147, 148, 150, 151 and 152	IMVS

*

a. Institute of Medical and Veterinary Science, Adelaide, Australia

b. 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark

c. 172 and 173 from Statens Serum Institut, Copenhagen, Denmark, the rest from IMVS

TABLE 2

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
20	*E. coli* type strains for O serotypes 153, 154, 155, 156, 157, 158 , 159 and 160	IMVS
21	*E. coli* type strains for O serotypes 161, 163, 164, 8, 9 and 124	IMVS
22	As pool #21, plus *E. coli* 0111 type strain Stoke W.	IMVS
23	As pool #21, plus *E. coli* 0111:H2 strain C1250-1991	See d
24	As pool #21, plus *E. coli* 0111:H12 strain C156-1989	See e
25	As pool #21, plus *S. enterica* serovar Adelaide	See f
26	*Y. pseudotuberculosis* strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See g
27	*S. boydii* strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See h
28	*S. enterica* strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65,:i:e,n,z,15 and 52:d:e,n,x,z15	IMVS

*

d. C1250-1991 from Statens Serum Institut, Copenhagen, Denmark
e. C156-1989 from Statens Serum Institut, Copenhagen, Denmark
f. *S. enterica* serovar Adelaide from IMVS
g. Dr S Aleksic of Institute of Hygiene, Germany
h. Dr J Lefebvre of Bacterial Identification Section, Laboratoroie de Santè Publique du Quèbec, Canada

TABLE 3

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
29	*E. coli* type strains for O serotypes 153, 154, 155, 156, 158, 159 and 160	IMVS
30	*E. coli* type strains for O serotypes 161, 163, 164, 8, 9, 111 and 124	IMVS
31	As pool #29, plus *E. coli* O157 type strain A2 (O157:H19)	IMVS
32	As pool #29, plus *E. coli* O157:H16 strain C475-89	See d
33	As pool #29, plus *E. coli* O157:H45 strain C727-89	See d
34	As pool #29, plus *E. coli* O157:H2 strain C252-94	See d
35	As pool #29, plus *E. coli* O157:H39 strain C258-94	See d
36	As pool #29, plus *E. coli* O157:H26	See e
37	As pool #29, plus *S. enterica* serovar Landau	See f
38	As pool #29, plus *Brucella abortus*	See g
39	As pool #29, plus *Y. enterocolitica* O9	See h
40	*Y. pseudotuberculosis* strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See i
41	*S. boydii* strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See j
42	*S. enterica* strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z15 and 52:d:e,n,x,z15	IMVS
43	*E. coli* type strains for O serotypes 1,2,3,4,10,18 and 29	IMVS
44	As pool #43, plus *E. coli* K-12 strains C600 and WG1	IVMS See k

*

d. O157 strains from Statens Serum Institut, Copenhagen, Denmark

e. O157:H26 from Dr R Brown of Royal Children's Hospital, Melbourne, Victoria

f. *S. enterica* serovar Landau from Dr M Poppoff of Institut Pasteur, Paris, France

g. *B. Abortus* from the culture collection of The University of Sydney, Sydney, Australia

h. *Y. enterocolitica* O9 from Dr. K. Bettelheim of Victorian Infectious Diseases Reference Laboratory Victoria, Australia.

i. Dr S Aleksic of Institute of Hygiene, Germany

J. Dr J Lefebvre of Bacterial Identification Section, Laboratoroie de Santè Publique du Quèbec, Canada

k. Strains C600 and WG1 from Dr. B.J. Backmann of Department of Biology, Yale University, USA.

TABLE 4 PCR assay result using primers based on the E. coli serotype O16 (strain K-12) O antigen gene cluster sequence

Gene	Function	Base positions of the gene	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
*rmlB**	TDP-rhamnose pathway	90-1175	#1064(91-109)	#1065(1175-1157)	1085bp	17	60°C
*rmlD**	TDP-rhamnose pathway	1175-2074	#1066(1175-1193)	#1067(2075-2058)	901bp	13	60°C
*rmlA**	TDP-rhamnose pathway	2132-3013	#1068(2131-2148)	#1069(3013-2995)	883bp	0	60°C
*rmlC**	TDP-rhamnose pathway	3013-3570	#1070(3012-3029)	#1071(3570-3551)	559bp	0	60°C
*gtf**	Galactofuranose pathway	4822-5925	#1074(4822-4840)	#1075(5925-5908)	1104bp	0	55°C
*wzx**	Flippase	3567-4814	#1072(3567-3586)	#1073(4814-4797)	1248bp	0	55°C
*wzy**	O polymerase	5925-7091	#1076(5925-5944)	#1077(7091-7074)	1167bp	0	60°C
*wbbI**	Galactofuranosyl transferase	7094-8086	#1078 (7094-7111)	#1079(8086-8069)	993bp	0	50°C
*wbbJ**	Acetyltransferase	8067-8654	#1080(8067-8084)	#1081(8654-8632)	588bp	0	60°C
*wbbK***	Glucosyl transferase	5770-6888	#1082(5770-5787)	#1083(6888-6871)	1119bp	0	55°C
*wbbL****	Rhamanosyltransferase	679-1437	#1084(679-697)	#1085(1473-1456)	795bp	0****	55°C

*, **, *** Base positions based on GenBank entry U09876, U03041 and L19537 respectively
**** 19 pools giving a band of wrong size

TABLE 5 PCR assay data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
wbdH	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
wzx	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	50°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0	50°C
wzy	9901-10953	#900(9976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0*	61°C
wbdL	10931-11824	#870(10931-10949)	#871(11824-11796)	894bp	7	60°C
wbdM	11821-12945	#868(11821-11844)	#869(12945-12924)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0**	65°C

* Giving a band of wrong size in all pools

** One pool giving a band of wrong size

TABLE 5A PCR specificity test data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 25-28) giving band of correct size	Annealing temperature of the PCR
wbdH	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0*	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
wzx	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	55°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0*	50°C
wzy	9901-10953	#900(9976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0**	60°C
wbdL	10931-11824	#870(10931-10949)	#871(11824-11796)	894bp	0	60°C
wbdM	11821-12945	#868(11821-11844)	#869(12945-12924)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0*	65°C

* 1 pool giving a band of wrong size

** 2 pools giving 3 bands of wrong sizes, 1 pool giving 2 bands of wrong sizes

TABLE 6 PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
wbdN	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0	55°C
			#1199(184-201)	#1200(531-514)	348	0	55°C
			#1201(310-327)	#1202(768-751)	459	0	55°C
wzy	O antigen	858-2042	#1203(858-875)	#1204(2042-2025)	1185	0*	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0	63°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
wbdO	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0**	62°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
wzx	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	O***	63°C
wbdP	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0	55°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
wbdR	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-13612)	369	0	55°C
			#1231(13384-13401)	#1232(13731-13714)	348	0	60°C

* 3 bands of wrong size in one pool, 1 band of wrong size in all other pools

** 3 bands of wrong sizes in 9 pools, 2 bands of wrong size in all other pools

*** 2 bands of wrong sizes in 2 pools, 1 band of wrong size in 7 pools

TABLE 6A PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 37-42) giving band of correct size	Annealing temperature of the PCR
wbdN	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0*	55°C
			#1199(184-201)	#1200(531-514)	348	0*	55°C
			#1201(310-327)	#1202(768-751)	459	0	61°C
wzy	O antigen polymerase	858-2042	#1203(858-875)	#1204(2042-2025)	1185	1**	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0***	60°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
wbdO	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0****	61°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
wzx	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	0	63°C
wbdP	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0*	60°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
wbdR	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-	369	0	50°C
			#1231(13384-13401)	#1232(13731-	348	0	60°C

* 1 band of wrong size in one pool

** pool #39 giving two bands, one band of correct size, the other band of wrong size in another pool.

*** 2 bands of wrong sizes in one pool

**** 3 bands of wrong sizes in 2 pools, 2 bands of wrong sizes in 2 other pools

TABLE 7

PCR assay data using primers based on the *Salmonella enterica* serotype C2 (strain M67) O antigen gene cluster sequence

Gene	Function	Base positions of the gene according to SEQ ID NO: 3	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Number of pools (out of 7) giving band of correct size	Annealing temperature of the PCR
wzx	Flippase	1019-2359	#1144(1019-1036)	#1145(1414-1397)	396bp	0	55°C
			#1146(1708-1725)	#1147(2170-2153)	463bp	0	55°C
			#1148(1938-1955)	#1149(2356-2339)	419bp	0	55°C
wbaR	Abequosyl transferase	2352-3314	#1150(2352-2369)	#1151(2759-2742)	408bp	0	55°C
			#1152(2601-2618)	#1153(3047-3030)	447bp	0	55°C
			#1154(2910-2927)	#1155(3311-3294)	402bp	0	55°C
wbaL	Acetyl transferase	3361-3875	#1156(3361-3378)	#1157(3759-3742)	399bp	0	55°C
			#1158(3578-3595)	#1159(3972-3955)	395bp	0	50°C
wbaQ	Rhamnosyl	3977-5020	#1160(3977-3994)	#1161(4378-4361)	402bp	0	55°C
			#1162(4167-4184)	#1163(4774-4757)	608bp	0	55°C
			#1164(4603-4620)	#1165(5017-5000)	415bp	0*	60°C
wzy	O polymerase	5114-6313	#1166(5114-5131)	#1167(5515-5498)	402bp	0**	55°C
			#1168(5664-5681)	#1169(6112-6095)	449bp	0	55°C
			#1170(5907-5924)	#1171(6310-6293)	404bp	0	55°C
wbaW	Mannosyl transferase	6313-7323	#1172(6313-6330)	#1173(6805-6788)	493bp	0	50°C
			#1174(6697-6714)	#1175(7068-7051)	372bp	0	55°C
			#1176(6905-6922)	#1177(7320-7303)	416bp	0	55°C
wbaZ	Mannosyl transferase	7310-8467	#1178(7310-7327)	#1179(7775-7758)	466bp	0	50°C
			#1180(7530-7547)	#1181(7907-7890)	378bp	0	55°C
			#1182(8007-8024)	#1183(8464-8447)	458bp	0	55°C

* Positive pool gives another band, which is also present in another pool. All other pools gave bands of wrong size.

** Band of wrong size in 6 other pools.

TABLE 8

PCR primers based on the *Salmonella enterica* serotype B (strain LT2) O antigen gene cluster sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 4	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Annealing temperature of the PCR
rmlB	TDP-rhamnose pathway	4099-5184	#1094 (4100-4117)	#1095(4499-4482)	400bp	55°C
rmlD	TDP-rhamnose pathway	5184-6083	#1092(5186-5203)	#1093(5543-5526)	358bp	50°C
rmlA	TDP-rhamnose pathway	6131-7009	#1090(6531-6548)	#1091(6837-6820)	308bp	55°C
rmlC	TDP-rhamnose pathway	7010-7561	#1088(7013-7030)	#1089(7372-7355)	360bp	55°C
ddhD	CDP-abequose pathway	7567-8559	#1112(7567-7584)	#1113(7970-7953)	404bp	55°C
ddhA	CDP-adequose pathway	8556-9329	#1114(8556-8573)	#1115(8975-8958)	420bp	60°C
ddhB	CDP-adequose pathway	9334-10413	#1116(9334-9351)	#1117(9816-9799)	483bp	45°C
ddhC	CDP-adequose pathway	10440-11753	#1118(10440-10457)	#1119(10871-10854)	432bp	60°C
abe	CDP-adequose pathway	11781-12680	#1100(12008-12025)	#1101(12388-12371)	381bp	55°C
wzx	Flippase	12762-14054	#1120(12762-12779)	#1121(13150-13133)	389bp	55°C
			#1122(12993-13010)	#1123(13417-13400)	425bp	55°C
			#1124(13635-13652)	#1125(14051-14034)	417bp	55°C
wbaV	Abequosyl transferase	14059-15060	#1126(14059-14076)	#1127(14421-14404)	363bp	45°C
			#1128(14688-14705)	#1129(15057-15040)	370bp	45°C
wbaU	Mannosyl transferase	15379-16440	#1130(15379-15396)	#1131(15768-15751)	390bp	60°C
			#1132(15850-15867)	#1133(16262-16245)	413bp	50°C
			#1134(16027-16044)	#1135(16437-16420)	411bp	60°C
wbaN	Rhamnosyl transferase	16441-17385	#1136(16441-16458)	#1137(16851-16834)	411bp	45°C
			#1138(16630-16647)	#1139(17087-17070)	458bp	55°C
			#1140(16978-16995)	#1141(17382-17365)	405bp	50°C
manC	GDP-mannose pathway	17386-18825	#1098(17457-17474)	#1099(18143-18126)	687bp	60°C
mabB	GDP-mannose pathway	18812-20245	#1096(18991-19008)	#1097(19345-19328)	355bp	55°C
wbaP	Galactosyl transferase	20317-21747	#1142(20389-20406)	#1143(20709-20692)	321bp	55°C

TABLE 9 PCR results using LT2 primers*

Strain name	O group	1094 - 1095	1092 - 1093	1090 - 1091	1088 - 1089	1112 - 1113	1114 - 1115	1116 - 1117	1118 - 1119	1100 - 1101	1120 - 1121	1122 - 1123	1124 - 1125	1126 - 1127	1128 - 1129	1130 - 1131	1132 - 1133	1134 - 1135	1136 - 1137	1138 - 1139	1140 - 1141	1098 - 1099	1096 - 1097	1142 - 1143
		rmlB	rmlD	rmlA	rmlC	ddhD	ddhA	ddhB	ddhC	abe	wzx	wzx	wzx	wbaV	wbaV	wbaU	wbaU	wbaU	wbaN	wbaN	wbaN	manC	manB	wbap
M8	A	y	y	y	y	y	y	y	y							y	y	y	y	y	y	y	y	y
P9603	B	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
M40	C1																							
M67	C2	y	y	y	y	y	y	y	y													y	y	y
M18	D1	y	y	y	y	y	y	y	y							y	y	y	y	y	y	y	y	y
M388	D2	y		y	y	y	y	y	y													y	y	y
M840	D3	y	y	y	y	y	y	y	y							y	y			y	y	y		y
M32	E1	y	y	y	y																	y	y	y
M324	F	y	y	y	y																			
M338	G																							
M252	H																							
M364	I																							
M354	J																							
M335	K																							
M7	L																							
M369	M	y	y	y																				
M370	N																							
M95	O																							
M360	P																							
M373	Q	y																						
M261	R																							
M282	S																							
M387	T	y	y	y	y																			
M295	U																							
M289	V																							
M296	W																							
M278	X																							
M298	Y																							
M93	Z																							
M291	51																							
M309	52																							
M303	53	y	y	y																				
M292	54	y	y	y	y																	y	y	y
M304	56	y																						
M293	57	y	y	y																				
M305	58																							
M285	59	y	y	y	y																			
M306	60	y																						
M328	61																							
M330	65																							
M332	66																							
M1408	55																							y
M1409	62																							
M1410	63	y																						
M1413	67	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
M74	E4	y	y	y	y																	y	y	y

* y indicates a positive PCR result. Blank indicates a negative result.

TABLE 10 Gene specificities in *Salmonella enterica* serogroups

Serogroup	Genes										
	wzy	wzx	wbaP	wbaU	wbaN	wbaV	wbaO	wbaW	wbaZ	wbaQ	wbaR
A	B	D	B	B	B	D	-	-	-	-	-
B	B	B	B	B	B	B	-	-	-	-	-
D1	B	D	B	B	B	D	-	-	-	-	-
D2	E1	D	B	-	E1	D	E1	-	-	-	-
D3	D3	D	B	B	B	D	-	-	-	-	-
C2	C2	C2	B	-	-	-	-	C2	C2	C2	C2
E1	E1	E1	B	-	E1	-	E1				

- means 'not present'